Exam #1 - sketched solution

1. (a) There are three types of photosensitive cells in the human retina called cones which are responsible for color perception. Each cell has a different sensitivity function $S_i(\lambda)$ which measures the influence of each wavelength on the cell response. The response of the i-th cone is

$$R_i = \int S_i(\lambda)\ S(\lambda)\ d\lambda$$

   where S is the spectrum of the electromagnetic radiation hitting the retina and λ is the wavelength.

   (b) colors are synthesized by adding different amounts of three primary colors with different spectra e.g., R, G, B;

   (c) one way of characterizing objects with multiple colors is by splitting the color space (RGB) into non-overlapping cells and computing the color histogram associated to the object.

2.

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} ax_2 \\ bx_1 + cx_2 \end{bmatrix} = \begin{bmatrix} x_2 & 0 & 0 \\ 0 & x_1 & x_2 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = M(x)\theta$$

where $\theta = [a\ b\ c]^T$. The energy is now

$$E = \sum_i \|y^i - M(x^i)\theta\|^2$$

Minimization of E: a necessary condition is

$$\frac{dE}{d\theta} = 0 \quad \Rightarrow \quad \frac{d}{d\theta}\sum_i (y^i - M(x^i)\theta)^T(y^i - M(x^i)\theta) = 2\sum_i M(x^i)^T(y^i - M(x^i)\theta) = 0$$

Therefore $\hat{\theta}$ must be a solution of the following system of equations

$$R\hat{\theta} = r$$

$$R = \sum_i M(x^i)^T M(x^i) \qquad r = \sum_i M(x^i)^T y^i$$

3. (a) x_1, x_3, x_7 are known; only the other 5 variables have to be estimated by minimizing S Let $\bar{S} = \{2, 4, 5, 6, 8\}$ be indices of the unknown variables. Then

$$\frac{dS}{dx_n} = 0\ ,\ \forall n \in \bar{S}$$

   These derivatives can be easily computed for middle points and they have a slightly different expression at the boundaries

$$\frac{dS}{dx_n} = \frac{d}{dx_n}(x_n - x_{n-1})^2 + (x_n - x_{n+1})^2 = 2(x_n - x_{n-1}) + 2(x_n - x_{n+1}) = 0\ ,\ \forall n \in \bar{S}\backslash\{8\}$$

$$\frac{dS}{dx_n} = \frac{d}{dx_n}(x_n - x_{n-1})^2 = 2(x_n - x_{n-1}) = 0\ ,\ n = 8$$

   Therefore, we have $x_1 = 1, x_3 = 3, x_7 = -1$ and the other variables verify

$$\begin{array}{llll}
-x_1 + 2x_2 - x_3 & = 0 & 2x_2 & = x_1 + x_3 \\
-x_3 + 2x_4 - x_5 & = 0 & 2x_4 - x_5 & = x_3 \\
-x_4 + 2x_5 - x_6 & = 0 & -x_4 + 2x_5 - x_6 & = 0 \\
-x_5 + 2x_6 - x_7 & = 0 & -x_5 + 2x_6 & = x_7 \\
-x_7 + x_8 & = 0 & x_8 & = x_7
\end{array}$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 2 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 3 \\ 0 \\ 0 \\ 0 \\ -1 \\ 0 \end{bmatrix}$$

Solution: $x = [1\ 2\ 3\ 2\ 1\ 0\ -1\ -1]$

(b) If the three observations are noisy we have to estimate the eight variables $x = [x_1, \ldots, x_8]^T$. This can be done by minimizing an energy functional with two terms: a data fit term and a regularization term

$$E(x) = \sum_{i \in \mathcal{S}} (y_i - x_i)^2 + \lambda \sum_{i=2}^{8} (x_i - x_{i-1})^2$$

The minimization of $E(x)$ leads to

$$\begin{bmatrix} 1+\lambda & -\lambda & 0 & 0 & 0 & 0 & 0 & 0 \\ -\lambda & 2\lambda & -\lambda & 0 & 0 & 0 & 0 & 0 \\ 0 & -\lambda & 1+2\lambda & -\lambda & 0 & 0 & 0 & 0 \\ 0 & 0 & -\lambda & 2\lambda & -\lambda & 0 & 0 & 0 \\ 0 & 0 & 0 & -\lambda & 2\lambda & -\lambda & 0 & 0 \\ 0 & 0 & 0 & 0 & -\lambda & 2\lambda & -\lambda & 0 \\ 0 & 0 & 0 & 0 & 0 & -\lambda & 1+2\lambda & -\lambda \\ 0 & 0 & 0 & 0 & 0 & 0 & -\lambda & \lambda \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 3 \\ 0 \\ 0 \\ 0 \\ -1 \\ 0 \end{bmatrix}$$

4. The gradient algorithm is a simple way to detect edge points in images. the gradient vector of an image $I(x, y)$ is the vector of partial derivatives of I with respect to x, y

$$g(x) = \begin{bmatrix} \frac{\partial I}{\partial x} \\ \frac{\partial I}{\partial u} \end{bmatrix}$$

and it is often characterized by its magnitude $\|g(x)\| = \sqrt{g_1(x)^2 + g_2(x)^2}$ and orientation $\theta = \arctan\left\{\frac{g_2(x)}{g_1(x)}\right\}$. In discrete images, the partial derivatives are replaced by finite differences or obtained by filtering the image with high pass filters (e.g., Sobel filters).

The gradient algorithm is based on three steps:

- gradient computation: g_1, g_2 can be obtained by convolving the image with the impulse response of the Sobel filters

$$\begin{array}{l} g_1(m,n) = h_1(m,n) * I(m,n) \\ g_2(m,n) = h_2(m,n) * I(m,n) \end{array} \qquad h_1(m,n) = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}, \qquad h_2(m,n) = h_1(n,m)$$

- thresholding: the gradient magnitude at each point is compared with a threshold $\|g(m,n)\| > T$. If the magnitude exceeds the threshold, the point (m,n) is accepted as an edge candidate.
- non-maximum suppression: we then check if the gradient magnitude has a local maximum at (m,n) along the direction of the gradient. The point is accepted as an edge point if it meets these two conditions.

5. (a) Suppose $X = X_i + \alpha V$ is a point of a straight line with direction V containing the point X_i. The point X is projected onto the camera plane on a point with coordinates

$$\tilde{x}_1 = \frac{\pi_1^T \tilde{X}}{\pi_3^T \tilde{X}} \qquad \tilde{x}_2 = \frac{\pi_2^T \tilde{X}}{\pi_3^T \tilde{X}}$$

Using the equation of the line we obtain

$$\tilde{x}_1 = \frac{\pi_1^T(\tilde{X}_i + \alpha\tilde{V})}{\pi_3^T(\tilde{X}_i + \alpha\tilde{V})} \qquad \tilde{x}_2 = \frac{\pi_2^T(\tilde{X}_i + \alpha\tilde{V})}{\pi_3^T(\tilde{X}_i + \alpha\tilde{V})}$$

where $\tilde{X} = \begin{bmatrix} X \\ 1 \end{bmatrix}, \tilde{V} = \begin{bmatrix} V \\ 0 \end{bmatrix}$. When α tends to infinity, the point X moves on the line towards infinity and the projection converges to

$$\tilde{x}_1 = \frac{\pi_1^T\tilde{V}}{\pi_3^T\tilde{V}} \qquad \tilde{x}_2 = \frac{\pi_2^T\tilde{V}}{\pi_3^T\tilde{V}}$$

These are the coordinates of the vanishing point, which do not depend on X_i. They depend only on the direction of the line.

(b) The previous equations define two restrictions on the camera parameters

$$\tilde{x}_1\pi_3^T\tilde{V} = \pi_1^T\tilde{V} \qquad \tilde{x}_2\pi_3^T\tilde{V} = \pi_2^T\tilde{V}$$

which can be written as follows

$$\begin{bmatrix} V^T & 0 & -x_1V^T \\ 0 & V^T & -x_2V^T \end{bmatrix} \begin{bmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

If we know M sets of parallel lines ($M >= 6$) we obtain a system of equations $M\pi = 0$ similar to the one of the linear calibration method. Therefore we could use in principle this information to estimate the camera matrix P. This is not true however since the projected points $\tilde{V}$ have the last coordinate equal to zero. This means that the last column of P does not influence the vanishing point and cannot be estimated in this way.